Urban Rail Transit and Endogenous Worker Choices*

Justin Tyndall†

September 30, 2022

Abstract

The provision of public transportation can improve the accessibility of work opportunities, particularly for low-income residents. However, predicting the labor market effects of new transit infrastructure is difficult because workers may endogenously adjust location, commute mode, and work decisions based on the existence of the infrastructure. I examine a large public transit rail project on the island of O'ahu, Hawai'i. Using a matrix of census block level travel times, I propose and estimate a structural model of location and mode choice for workers on O'ahu. I use estimated parameters to predict the effects of the rail project. I find the new rail system substantially increases the use of public transit and increases the employment rate, particularly among low-income workers. Accounting for endogenous household decisions through structural modeling is key to the findings. High-income workers have a strong preference for driving and a low preference for the neighborhoods in which rail was constructed, this allows low-income workers to capture a high share of transit's direct benefits.

Keywords: Transportation, Transit, Residential Choice, Neighborhood Change, Spatial Mismatch
JEL classification: J20, J60, R13. R23, R40, R58

*Thank you to the W.E. Upjohn Institute for Employment Research for their support through the Early Career Research Grant program.

†University of Hawai'i Economic Research Organization and Department of Economics, University of Hawai'i at Manoa, Honolulu, HI, United States. *jtyndall@hawaii.edu*.

1 Introduction

Constructing public transit infrastructure can improve labor market opportunities by reducing commuting costs. However, estimating the commuter benefits of new transit infrastructure is challenging due to endogenous worker responses and land market effects. Workers may change their home location, work location, or labor market participation in response to new transit infrastructure, and all of these decisions will affect the magnitude and distribution of transit's benefits.

I study the implementation of rail transit on O'ahu, Hawai'i. Prior to the recent rail project, O'ahu had no rail transportation. The rail line constructed was meant to benefit commuters by providing an alternative to a major commuting artery, which suffers from heavy rush hour vehicle traffic. While the system will provide large time savings relative to the existing bus service, the long-run effects are unknown without accounting for endogenous worker decisions. I collect detailed, block level commute time data and block level bilateral commuter flow data. Through a structural estimation approach, I estimate worker preferences across commuting routes and modes, for both low and high wage workers. I then apply these parameters to estimate the general equilibrium effects of the new rail infrastructure on commute times, public transit mode share, and employment. Under static worker choice, I find rail produces commute time savings for the average worker. After accounting for endogenous decisions, I find the rail system leads to a small *increase* in the average commuting time on O'ahu, as workers substitute away from cars and towards transit, and substitute towards routes for which they have a strong preference but had previously avoided due to high commute costs. Despite failing to reduce average commute time in the long run, I find the rail system leads to significant increases in public transit mode share and a significant increase in the aggregate employment rate.

Longitudinal data tracking individual's changes in home and work location is not typically available to researchers analyzing the effects of transportation systems. As a result, understanding endogenous location decisions typically relies on directly modeling the choices of workers. Structural neighborhood models have been implemented to estimate aggregate and distributional benefits of new urban amenities, particularly transportation systems. The basis for spatial urban models comes from the monocentric city model (Alonso, 1964; Muth, 1969; Mills, 1967), and polycentric city model (Fujita and Ogawa, 1982). Workers accept higher commuting time to access areas with lower housing costs. In a spatial equilibrium, these costs and benefits must lead to an equalization of utility over space. The extension of the basic urban model to incorporate structural modeling approaches, based on the discrete choice methods of McFadden (1973), was developed in Anas (1981) and Epple and Sieg (1999) and further extended in several papers including Bayer et al. (2004), Sieg et al. (2004), Bayer et al. (2007), Bayer and McMillan (2012), and Ahlfeldt et al. (2015).

This paper relates most closely to a recent literature on estimating benefits of transit infrastructure using structural neighborhood choice modeling. Severen (2019) examined the impact of rail transit on the labor market in Los Angeles through a structural estimation approach. Tsivanidis (2018) provided an analysis of a Bus Rapid Transit (BRT) system on the labor market in Bogota. Tyndall (2021) analyzed the effect of Light Rail Transit (LRT) on labor markets across four US cities. Chernoff and Craig (2022) examined distributional effects of a rail expansion in Vancouver. Each of these papers implemented a structural neighbhorhood choice model to understand the interaction between housing markets, labor markets and endogenous worker decisions in estimating the effects of transit infrastructure. I incorporate features of these models to develop a simple framework that can identify the aggregate effects of rail. I apply the model to a data set with more spatial detail than has been used in past literature. The richness of data and the unique setting allow for a clear identification of system impacts.

The theory that spatial access to jobs may drive joblessness was proposed as the spatial mismatch hypothesis in Kain (1968). Andersson et al. (2018) provided important empirical work that confirmed the continued importance of spatial labor market access to worker's job prospects in the US. Some work has looked specifically at public transit and the role of transit access in labor market outcomes. Sanchez (1999) demonstrated an empirical connection between low transit access and joblessness. Some papers have relied on natural experiments where transit access changed exogenously to identify causal labor market effects (Holzer et al., 2003; Tyndall, 2017), these studies found a positive impact of transit access on employment.

A specific focus of this paper is to understand the role of long-run endogenous sorting on the impacts of new rail infrastructure. By executing a model across several stages of a rail phase-in period, I estimate the relative role of direct commuting cost reductions and the role of endogenous household location, mode, and work decisions. I specifically calculate effects on average commuting time, transit mode share and the island-wide employment rate. I find that only accounting for direct commuting cost savings fails to capture the aggregate impact of transit. Workers with strong preferences for using transit are likely to sort towards stations, while workers with a preference for driving will sort away from stations, repelled by rising land costs. Low-wage workers are more likely to use transit, but are also more sensitive to rent increases, meaning the effect of a local public transit amenity that raises neighborhood demand might attract, or repel low-wage workers, depending on the magnitude of the two effects (Tyndall, 2021). The structural approach attempts to account for these competing effects and estimate the total island-wide impacts.

The paper will proceed as follows. Section 2 describes the empirical setting. Section 3 provides a discussion of data. Section 4 describes the structural estimation methodology. Section 5 provides results and Section 6 concludes.

2 Empirical Setting

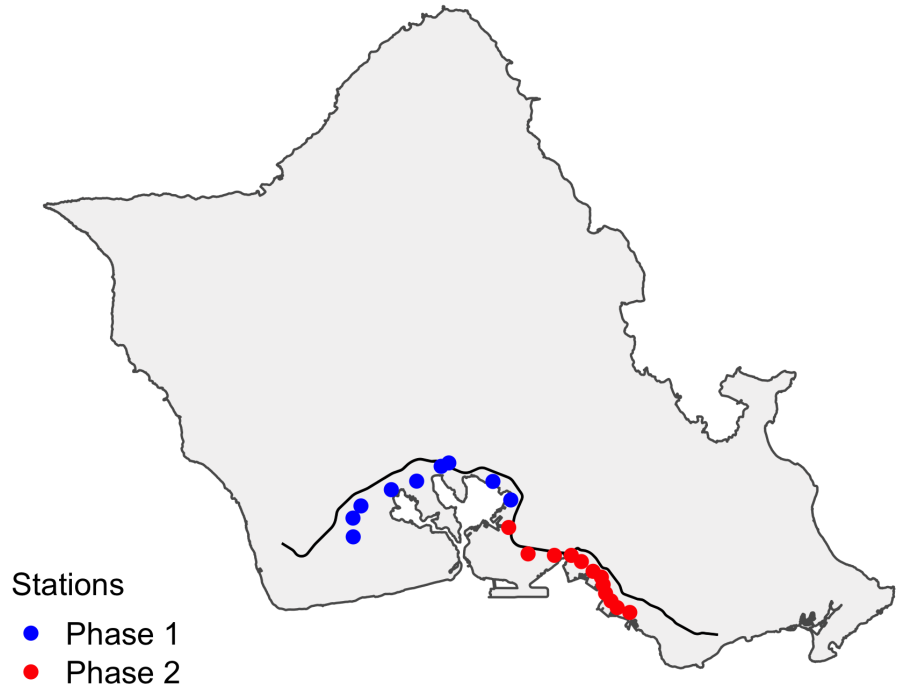
I study the construction O'ahu's first public transit rail line. The system has a so-called "hybrid-rail" design, combining features of both light and heavy rail systems. The system is elevated, with track and station platforms supported on concrete pillars. The full line is planned to include 21 stations, which span 31 km. The western edge of the system extends to the Kapolei neighborhood, and the easternmost station is located at Ala Moana Center, a major shopping area in the urban core of Honolulu.[1] The opening of the full 21 station line is set to be completed in two stages, with the westernmost nine stations opening in 2023 and the remaining 12 stations scheduled to open in 2031. I will refer to these two sections of the line as Phase 1 and Phase 2 respectively, and provide analysis on the effects of both Phase 1 as well as the full line (Phase 2). Figure 1 shows the locations of the rail stations on the island of O'ahu.

The path of the rail line roughly follows the H1 Interstate Highway. The H1 serves commuters from the west side of the island who commute into the urban core of Honolulu. East-bound traffic on the H1 is severe during rush hour, which served as a partial motivation for providing a public transit option on this route. Household incomes on the west side of O'ahu are generally lower than on the east side of O'ahu, meaning the proposed rail route is aligned to provide access to the downtown job center for working class populations.

The history of passenger rail planning on O'ahu spans several decades. City documents discussing the prospect of an urban rail line can be found dating back to 1967. In 2005, funding was secured to begin construction of the project and in 2011 construc-

[1]The precise location of the easternmost stations are the topic of policy debate and could be revised.

Figure 1: Location of Rail Stations on O'ahu and the H1 Highway



The H1 Interstate Highway is shown in black. Phase 1 stations are scheduled to open in 2023. Phase 2 stations are scheduled to open in 2031.

tion began. The rail project has experienced significant delays in construction and large cost overruns. Even after construction began, there was significant political uncertainty regarding whether the project would be completed. For example, mayoral campaigns since 2005 have centered on whether to complete or abandon construction of the rail line. Political opposition to the construction of rail often centered on concerns with cost overruns. When construction began, capital costs were expected to be $3 billion, with $1.6 billion coming from the Federal Transit Administration (FTA). Projected costs rose steadily over the following years. The current projected cost of the line is $12.4 billion. Even considering the high costs of transportation infrastructure throughout the US (Brooks and Liscow, 2022; Gupta et al., 2022), the O'ahu system construction costs are extremely high relative to comparable cities, in terms of either total cost or costs per system mile.

After the system is completed, the rail line will provide significantly improved public transit commuting for a large section of the island. Currently, O'ahu provides relatively extensive bus service compared to similar sized US cities. However, buses travel within general traffic in almost all cases, meaning they are subject to traffic delays and accompanying trip duration uncertainty. Public transit users along the line will expect to experience reduced commuting costs.

The island of O'ahu is coterminous with the City and County of Honolulu.[2] O'ahu provides an excellent study location for several reasons. First, as a small island, the relevant local labor market is cleanly defined. Typically, studies of urban labor markets impose assumptions to define a study area, often adopting Census boundaries. In the case of O'ahu, the boundaries of the study area are clear and there are no border area spillover effects to be considered. Access to O'ahu from the neighboring Hawaiian islands is only possible by air travel. O'ahu is small enough that commuting is possible across the entire island, though large enough to be comparable in size to the commuting sheds of other US metropolitan areas. Second, the O'ahu rail system is a significant infrastructure investment and the first rail connection on the island. The lack of existing rail infrastructure makes the treatment definitions clearer, as I do not need to consider network effects for a pre-existing rail system.

O'ahu shares many design characteristics with mid-sized American cities, such as significant highway infrastructure and primarily single-family zoned land use, surrounding a relatively dense urban core. Demographics on O'ahu are unique in several dimensions. Median household income on O'ahu ($87,700) is higher than the median household income across US metropolitan areas ($69,600), while the college education rate is very similar. O'ahu has a high Asian population share (43%) and a high share of Native Hawaiians and Pacific Islanders (10%), when compared to other metros in the US. The odds that a worker commutes by public transit on O'ahu (7.2%) is about 40% higher than the rate across other metros. Demographic information for the study area is provided in Table 1, with comparisons to average US metro conditions and the US as a whole.

[2]Counties in Hawai'i do not contain distinct municipalities, rather they operate under a combined city-county system.

Table 1: Demographic Characteristics of Study Area

	O'ahu	**US Metros**	**USA**
Population	979,682	284,298,061	331,449,281
Median household income ($)	87,722	69,591	64,994
College education rate† (%)	35.7	34.7	32.9
Labor force participation (%)	66.4	64.3	63.4
Unemployment (%)	2.6	3.5	3.4
Median age	38.2	38.0	38.2
Owner-occupancy rate (%)	57.5	63.0	64.4
White (%)	20.2	68.2	70.4
Black (%)	2.5	13.4	12.6
Asian (%)	42.6	6.3	5.6
Native Hawaiian or Pacific Islander (%)	10.0	0.2	0.2
Hispanic (%)	10.0	20.6	18.2
Average commute time (minutes)	28.0	27.5	27.0
Commuter mode share:			
Drove alone (%)	78.6	83.2	83.8
Public transportation (%)	7.2	5.2	4.8
Walking (%)	5.6	2.5	2.6

Data are from the 2020 5-year American Community Survey.
† Bachelor's degree or above, among population 25 years and older.

3 Data

I construct a route level data set, with granularity at the census block level. I rely on block level bilateral commuting flow data from the 2015-2019 Longitudinal Employer-Household Dynamics Origin-Destination Employment Statistics (LODES) and a block level commuting time matrix provided by the transportation routing firm Travel Time.

LODES breaks out commuter flows by worker income. I categorize workers into two worker types, low and high-wage workers, relying on the cut-off values used in LODES. Low-wage workers are defined as those earning less than $39,996 annually and high-wage workers earn more than this amount. Across the 2015-2019 LODES, I observe 1,360,158 unique block-to-block commutes. Low-wage workers cover 956,121 unique routes, while high wage workers cover 755,136 unique routes. I treat these commutes as the possible commute choices available to each worker of that type. The routes include 4,960 unique home locations and 3,212 unique work locations.

I gather extensive trip level data from the transportation routing firm Travel Time. For any pair of latitude and longitude coordinates, the Travel Time Application Programming Interface (API) returned an estimate of the commuting time. I queried the

API for each travel route in the choice set. The API incorporates predicted traffic and transit schedule conditions for a selected time. I set parameters to collect data for the quickest possible route that would allow the worker to get to their destination by 9:00 am on a Wednesday in order to match likely commuting time. I use the geographic centroid of each census block as the origin and destination point, and calculate driving and transit times for all 1,360,158 observed routes. Figure 2 shows the relationship between driving times and public transit times for the data covering the period before the rail system was running. For every route, driving provides a shorter trip time than public transit. For 96.3% of routes, public transit takes more than twice as long as driving, for 74.7% of routes transit takes more than three times as long, and for 48.8% of routes transit takes more than four times as long. For the average route, drive time is 19.2 minutes while public transit time is 59.9 minutes. When weighted by the number of workers who take each route across the LODES data, I find average driving time to be 19.5 minutes, while average transit time for the same set of routes would be 54.7 minutes. One-way commute times and distances are reported in Table 2. The average commuting times calculated with Travel Time data are comparable to estimates from the American Community Survey (ACS).

Table 2: Summary Statistics, Route Level Data

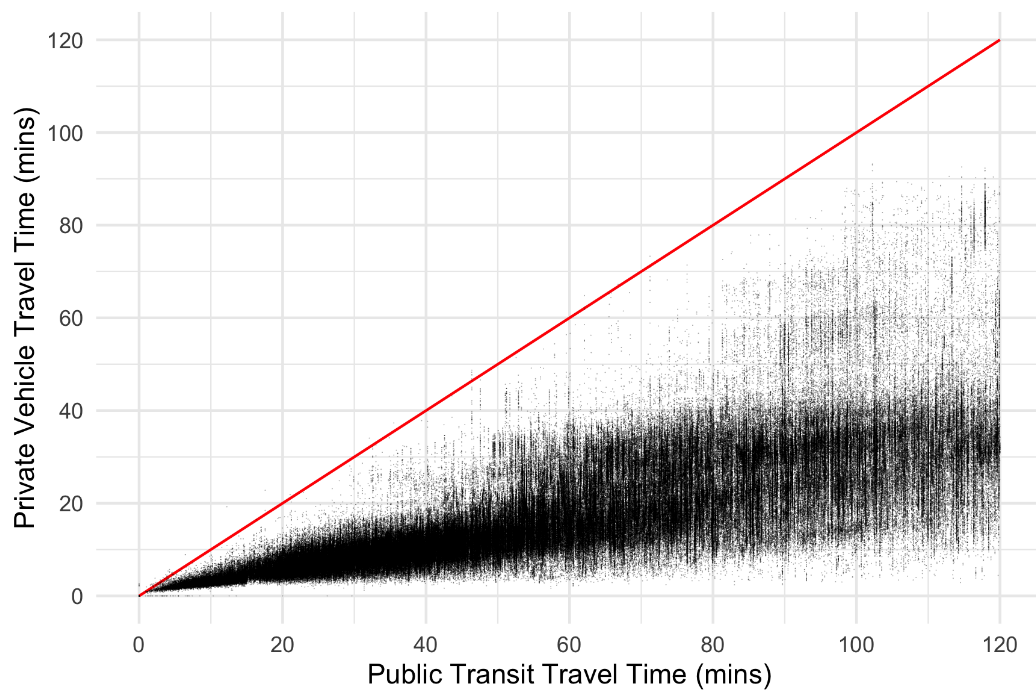
	All Observed Routes		Weighted by Workers	
	Driving	Transit	Driving	Transit
Average time (mins)	19.2	59.9	19.5	54.7
Average road distance (kms)	18.1	.	15.0	.

Average route characteristics among observed commutes on O'ahu. Public transit figures ignore routes that cannot be completed by transit or would take more than 2-hours one-way.

Figure 3 provides examples of the trip time data, showing the area that can be covered by driving and public transit for an example origin location. The left images show the area that can be covered within 30 minutes, while the right images show the area that can be covered in one hour. Comparing the top and bottom panels, the area accessible by driving in a given time is drastically larger than the area that can be accessed by public transit. Almost the entire island is accessible in a one hour drive, while only a small fraction is accessible through a one hour public transit commute. The figures reflect pre-rail commute times.

I restrict the data set by dropping any commute that is estimated to take more than

Figure 2: Drive Times vs Public Transit Times for Observed Commute Routes, Before Rail

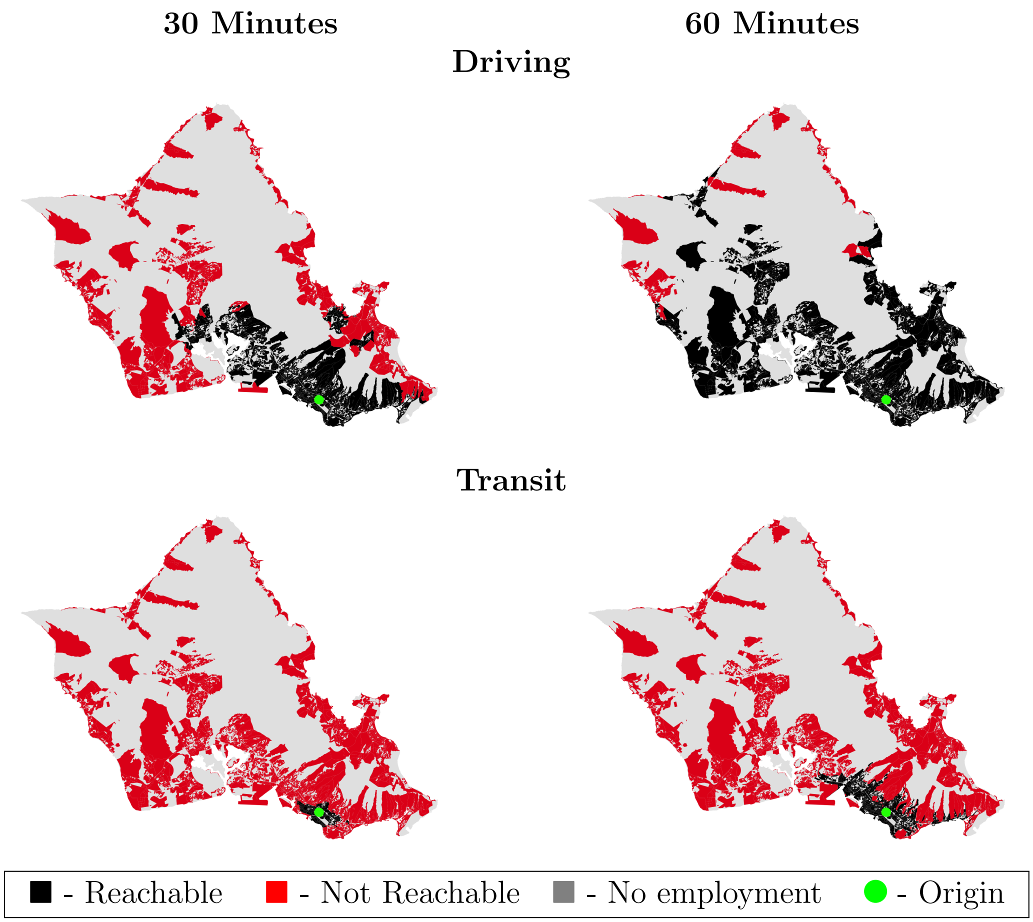


Each point represents one commuting route. The red line would indicate trips where private vehicle and transit commute times are equal. The figure displays all routes that can be completed in under two hours by both driving and public transit (744,968 observations).

two hours, one-way, as these are unlikely to be viable daily commutes. This restriction applies only to public transit commuting as there are no two census blocks on O'ahu that are more than two hours apart by driving.

The model will incorporate estimates of local housing costs as a parameter. I approximate annualized local housing costs for each census block in the model. I use Multiple Listings Service (MLS) data for O'ahu. The data covers every real estate transaction in the MLS from 2010-2021. I calculate the median sales price of a home at the census tract level, assume an annual price to rent ratio of 20, and assign annual housing costs to each block based on which tract it is located in. I estimate costs at the tract rather than block level to reduce noise in areas with few transactions. Estimated annual housing costs calculated in this way range from \$11,750 to \$137,397, with a median value of \$34,975 (or \$2,915 per month). Because the model encompasses both renters and owners, this method gives a more accurate approximation of spatial

Figure 3: Job Locations Accessible from One Origin Location



Block level information is presented, showing which areas are reachable through driving and public transit from an origin location placed in Honolulu's city-center. I find driving provides dramatically more job opportunities to a worker when compared to using public transit. The displayed data captures the pre-rail period.

variation in housing costs as compared to survey data on rents.

The model introduced below will also incorporate basic demographic information, such as the employment rate. For this demographic information I use 2020 5-year ACS.

I structure the data at the route level. I compile a menu of possible choices available to workers in the model. Workers choose among all routes that are observed with a positive frequency in the data for their worker type, and select either driving or public transit.

4 Methodology

I propose a structural neighborhood choice model to estimate the effects of the new rail system. With detailed commuter flow and travel time information I am able to explicitly model worker choices. I allow workers to choose their home location, work location, commute mode (driving vs transit), and labor market participation. The model is built on the assumptions of the classic urban monocentric city model. Under the monocentric model, workers are utility maximizing and face a trade-off between housing costs and commuting costs. Solving the model will yield commuting cost parameters over routes and modes and allow worker behavior to be estimated in counterfactual scenarios. The introduction of rail reduces some commuting costs. By holding constant worker preference parameters and resolving the model under alternative transit counterfactuals I am able to estimate the impact of rail on aggregate worker outcomes inclusive of endogenous worker decision making.

Equation 1 is a Cobb-Douglas style utility function which governs worker preferences.

$$U_{ijkv} = (C - c_{s(i)jkv})^{\gamma} H^{(1-\gamma)} \chi_{s(i)jk} \xi_{ijkv} \tag{1}$$

Workers derive utility from numeraire consumption (C) and the consumption of generic units of housing (H). Non-monetary commuting costs (c) reduce consumption utility. Each worker (i) chooses a home location (j), work location (k), and mode of transportation (v). Mode choice is limited to driving or public transportation. The share of income a worker spends on housing is set by $1 - \gamma$. Each worker is either a high ($s(i) = h$) or low ($s(i) = l$) wage worker. $s(i)$ determines the income level of the worker when they are employed and this characteristic is fixed.

$\chi_{s(i)jk}$ is a route and worker type specific preference parameter. Beyond differences in commuting costs (which are accounted for directly) some routes may provide higher utility than others based on their unique characteristics such as housing and job prospects or any other route specific characteristics. Given spatial differences in job types and housing quality, different worker types may have different common preferences. All workers of the same type share a common evaluation of χ_{jk}. Resolving $\chi_{s(i)jk}$ will help produce realistic substitution patterns in the counterfactuals as workers of specific types will preferentially substitute towards routes that provide higher utility to their type on average. A Type 1 extreme value distributed error term (ξ_{ijkv}) captures the worker specific idiosyncratic preferences over each available route-mode option.

Non-monetary commuting costs ($c_{s(i)jkv}$) are defined in Equation 2. $\zeta_{s(i)v}$ is the mode specific cost of commuting per hour as a share a worker's wage. $\zeta_{s(i)v}$ is allowed to differ across worker types, as they may have different preferences across modes. $\omega_{s(i)k}$ denotes hourly wage. τ_{jkv} represents the annual hours spent in commute.

$$c_{s(i)jkv} = \zeta_{s(i)v}\omega_{s(i)k}\tau_{jkv} \tag{2}$$

Each worker operates under a budget constraint, represented by equation 3. Workers exhaust their income[3] ($w_{s(i)k}$) on housing costs (Hp_j), numeraire consumption (C), and monetary commuting costs (θ_{jkv}). Monetary commuting costs will be calculated according to the mode selected and the distance of the commute. Workers choose a utility maximizing quantity of housing and pay the market housing costs in their home location (p_j).

$$w_{s(i)k} = Hp_j + C + \theta_{jkv} \tag{3}$$

A worker's income is set according to their type, except in the case where a worker chooses a null work location ($k = \emptyset$), which represents being out of the labor force. When out of the labor force a worker pays no commuting costs, and receives a government allocated income (ι).

The utility function and budget constraint combine to produce an indirect utility function, show in Equation 4.

$$V_{ijkv} = (w_{s(i)k} - c_{s(i)jkv} - \theta_{jkv})\gamma^{\gamma}\frac{1-\gamma}{p_j}^{1-\gamma}\chi_{s(i)jk}\xi_{ijkv} \tag{4}$$

The extreme value distributed idiosynncratic error term produces a multinomial logit probability function (Equation 5), capturing the probability a worker selects a specific home, work, mode triple (P_{ijkv}). Upper bar notation indicates the maximum value in the set.

$$P_{ijkv} = \frac{e^{V_{ijkv}}}{\sum_1^{\bar{j}}\sum_1^{\bar{k}}\sum_1^{\bar{v}} e^{V_{ijkv}}} \tag{5}$$

I calculate the total public transit mode share by summing all of the choice prob-

[3] Annual income ($w_{s(i)k}$) and hourly wage ($\omega_{s(i)k}$) are related assuming an eight hour work day and 260 working days in a year: $w_{s(i)k} = \omega_{s(i)k} \times 8 \times 260$.

abilities where v is public transit and dividing by the total sum of choice probabilities. I will refer to the true (observed) public transit mode share as $M_{s(i)}$ and the model generated value as $\mathcal{M}_{s(i)}$.

$$\mathcal{M}_{s(i)} = \frac{\sum_1^{\overline{j}} \sum_1^{\overline{k}} P_{ijk(v=\text{transit})}}{\sum_1^{\overline{j}} \sum_1^{\overline{k}} \sum_1^{\overline{v}} P_{ijkv}} \tag{6}$$

5 Solution Method

I first solve the complete model using data from the pre-rail period. I make use of cross-sectional variation in worker commuting behavior to recover preference parameters governing commute time, mode choice, and a vector of route and worker type preference parameters. I then use these parameters to run four counterfactual scenarios, which I refer to as Scenarios 2-5. These subsequent scenarios capture conditions across various rail and worker sorting conditions as described below.

To estimate the model, I impose several exogenous parameters, shown in Table 3. Annual income is set to \$19,859 for low-wage workers and \$85,326 for high-wage workers. I recover these estimates from ACS microdata.[4] I set the out of labor force income to be \$10,000.

Table 3: Exogenous Model Parameters

Symbol	Value	Description
$w_{s(i)=l}$	19.859	Low-wage worker income (\$1,000)
$w_{s(i)=h}$	85.623	High-wage worker income (\$1,000)
ι	10.000	Out of labor force income (\$1,000)
γ	0.67	Share of income spent on non-housing consumption
$M_{s(i)=l}$	0.180	Initial public transit mode share, low-wage workers
$M_{s(i)=h}$	0.085	Initial public transit mode share, high-wage workers
$\zeta_{v=\text{driving}}$	0.93	Commuting cost per unit time as share of wage, driving
$\theta_{jk(v=\text{transit})}$	0.96	Annual monetary cost of transit commuting (\$1,000)
$\theta_{jk(v=\text{driving})}$	$0.0589 \times \mathrm{d}_{jk}$	Annual monetary cost of private vehicle commuting (\$1,000), d=distance in km

I impose these parameters on the model.

I assume workers spend a constant fraction of income on housing $(1-\gamma)$. Davis and Ortalo-Magné (2011) discuss and estimate this parameter for the US, finding that

[4] I use individual wage earnings from the 2020 5-year ACS microdata for Honolulu County. I drop workers with earnings of zero or less and take the mean value for workers in each wage category (low vs high). I find that the main results are not sensitive to moderate changes in wage level assumptions.

the average worker spends 24% of their income on housing. Using Honolulu County specific census microdata from the 2020 5-year ACS, I calculate the share of household income spent on gross rent or mortgage payments to be 33%, for the average household. The higher value reflects high housing costs in Honolulu. I therefore adopt $\gamma = .67$ in estimation.

I impose an estimate of the time cost of driving as a share of the wage rate. I select the parameter estimated in Small et al. (2005), which examined commuting behavior in Los Angeles, finding drivers faced a time cost of driving equal to 93% of their wage rate.

I constrain the model to produce the public transit mode share observed in aggregate data. I impose mode share restrictions that are specific to worker type. I identify $M_{s(i)}$ directly from ACS data as 18.0% for low-wage workers and 8.5% for high-wage workers. I consider walking to be a component of public transit to avoid introducing an additional mode choice. Notably, public transit mode share among the low-wage group is more than twice that of the high-wage group. When solving the model, the worker type specific time costs of public transit use ($\zeta_{s(i)v=\text{transit}}$) are determined endogenously and allow the model to generate the correct public transit mode shares in the pre-rail Scenario.

I impose monetary commuting costs (θ_{jkv}). For public transit users, I assume workers pay for 12 monthly transit passes each year, which costs \$960 in Honolulu ($\theta_{jk(v=\text{transit})} = 960$). For those driving, I approximate monetary commute costs using data from the American Automobile Assosiation (AAA) (American Automobile Association, 2021). Assuming 260 working days in a year, AAA estimates of marginal commuting costs for a "medium sedan" imply \$58.87 in annual costs for every km of daily commuting. For each route I use the driving distance estimated in the Travel Time data. To arrive at route specific monetary costs, I multiply the two-way commute distance by the per km cost of driving. Weighted by commuter frequency, I estimate the average cost of commuting by vehicle to be \$1,769 per year. I assume workers ignore the fixed costs of car ownership when choosing a commuting mode, as the decision to own a car reflects general mobility demand beyond commuting.

Workers implicitly make a labor force participation decision, as selecting a null work location ($k = \emptyset$) represents not working. When calculating the worker shares for $k = \emptyset$ "routes," I use ACS data on the number of residents in each census tract who are out of the labor force, and spread these workers uniformly across the tract's constituent blocks, as weighted by block population. I then scale up the number of workers out of labor force to precisely match the island-wide labor force participation rate as recorded in the ACS data (66.4%). I assume worker non-participation is equally likely across worker types.

The model is solved through contraction mapping, allowing workers to iteratively select home locations, work locations, and mode. Route level idiosyncratic parameters adjust to attract the precise number of workers of each wage type as is recorded in the data. Furthermore, $\zeta_{s(i)v=\text{transit}}$ parameters for the time cost of transit commuting are adjusted endogenously to ensure $\mathcal{M}_{s(i)} = M_{s(i)}$ for each $s(i)$. I define an equilibrium as the case where low and high wage worker flows precisely match the observed data, worker-type specific transit mode shares are matched to the data, and workers are in a Nash Equilibrium where they cannot improve their utility by altering any of their home, work, or mode decisions.

The model is identified through matching the observed commuter flows of the 1,720,907 route-type choices available by adjusting a vector of 1,720,907 endogenous route-type preference parameters ($\chi_{s(i)jk}$), and matching the two transit mode share values ($M_{s(i)}$) by adjusting a vector of two endogenous transit time cost parameters ($\zeta_{s(i)v=\text{transit}}$).

When solving the model I identify a unique equilibrium point. The uniqueness of the solution follows from Brouwer's fixed-point theorem. Bayer and Timmins (2005) discussed establishing uniqueness specifically for spatial sorting models. When neighborhood preference is partially determined by the characteristics of other members of the neighborhood (eg preference for neighbor income or race), multiple equilibria will naturally become a problem. In the current model, I do not consider neighbor preference, which removes concerns over the possible presence of multiple equilibria.

Identification of parameters in the pre-rail period (Scenario 1) comes from cross-sectional variation in worker choice. Holding other factors constant, if two routes in the model provide the same commute times, the routes will be chosen with equal frequency. To the extent workers in the data prefer one route over the other, the shared idiosyncratic preference parameter is raised to capture any characteristics of the route that might explain its relative popularity. An identifying assumption is that these preference parameters over routes remain fixed, and what changes is the matrix of public transit commute times. A reduction in public transit commute time makes a worker marginally more likely to prefer that route. Notably, preferences over home locations and work locations that may differ across worker types are nested within the route level preference parameters.

After solving Scenario 1, I estimate conditions under counterfactual scenarios. In Scenario 2, I lock-in the route level preference parameters and the preference parameters governing transit time costs and I adjust the matrix of public transit commute times to approximate the new rail service. I identify all commute routes for which a straight line connecting the origin and destination passes through the corridor where rail service will be available. I identify this corridor as a 3 km buffer area surrounding system stations. For each intersecting route I reduce the public transit commuting time by 30%.[5] I then recalculate worker commuting times under the improved public transit conditions, holding worker behavior fixed (Scenario 2). Subsequently, I allow workers to adjust behavior and use contraction mapping to solve for the new equilibrium under the new commute time matrix. Because some neighborhoods may experience an excess demand by residents, I allow local housing costs (p_j) to endogenously adjust so that the number of residents in each block is held constant. Offered wages are held constant, but I allow firms to endogenously shrink or grow if they experience a change in labor demand by workers. The new solution corresponds to Scenario 3. I calculate Scenario 4 and 5 solutions similarly. I first reduce public transit travel times for routes that intersect the Phase 2 rail area but did not intersect Phase 1 and recalculate commuting times holding worker behavior fixed at Scenario 3 levels. Scenario 5 solves the model for a third time through contraction mapping, considering the effects of the full rail system. The scenarios are summarized in Table 4. Providing estimates across the five scenarios is meant to highlight the role of endogenous worker choice, contrast these effects with those under static worker assumptions, and to roughly correspond to the chronological progression of rail construction and worker sorting.

6 Results

Solving the model under the pre-rail period (Scenario 1) yields the needed parameter values. I solve for the 1.7 million preference parameters covering all possible route choices ($\chi_{s(i)jk}$). Simultaneously, I recover the parameters for the time cost of using public transit as a share of wage ($\zeta_{s(i)v=\text{transit}}$), for both low and high wage workers. I estimate public transit time cost parameters of 2.21 and 1.20 for low and high wage workers respectively. The imposed time cost parameter of driving was 0.93. The estimated parameters suggest that a minute of public transit commuting is significantly

[5] In future versions of this paper I will replace this assumption by directly incorporating an updated public transit commuting time matrix from Travel Time.

Figure 4: Estimation Scenarios

Scenario 1	Pre-rail.
Scenario 2	Phase 1 rail is completed. Worker choices are held constant at Scenario 1 level.
Scenario 3	Phase 1 rail is completed. Endogenous worker choices.
Scenario 4	Phase 2 rail is completed. Worker choices are held constant at Scenario 3 level.
Scenario 5	Phase 2 rail is completed. Endogenous worker choices.

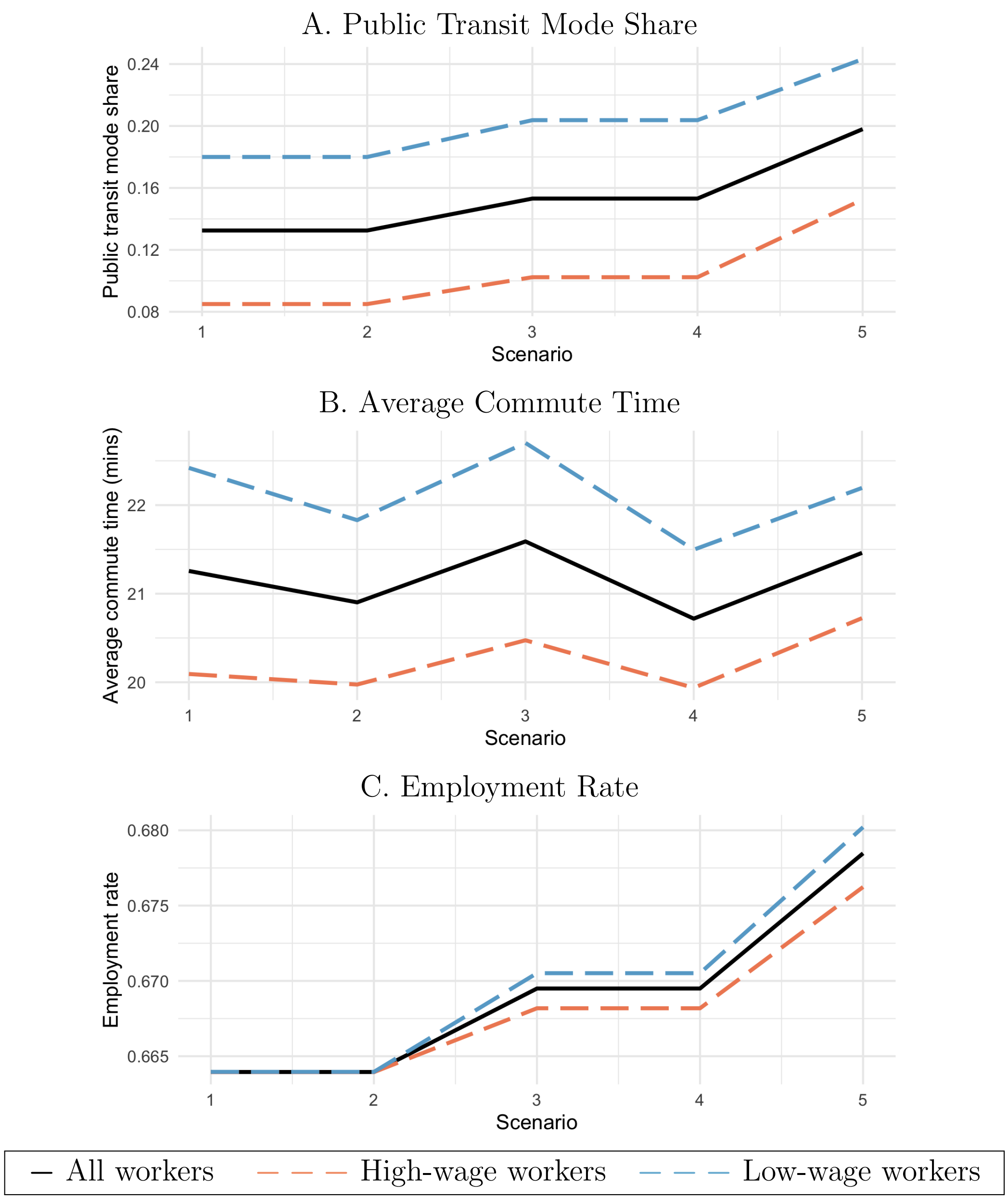
A description of the scenarios estimated. The locations of Phase 1 and Phase 2 rail stations are shown in Figure 1.

more costly than a minute of private vehicle commuting. The higher costs of taking transit are consistent with the inconveniences of public transit. For example, costs associated with adhering to a rigid schedule, and the possibility of missing a bus have been shown to discourage transit use (Tyndall, 2018). Considering the incomes of low and high wage workers, the parameters imply that a low-wage worker would have to be paid \$21.08 to be indifferent to enduring one additional hour of commuting by transit, while a high-wage worker would need to be paid \$49.53. These endogenous parameters are held constant through the five scenarios and I provide estimates for the counterfactual scenarios below.

I am primarily interested in estimating the effect of the rail system on commute times, public transit mode share, and the employment rate. I summarize the three outcomes across Scenarios in Figure 5.

Figure 5A shows the progression of public transit mode share across scenarios. In the pre-rail period, the model matches transit mode share to observed data, with 18.0% of low-wage workers using transit and 8.5% of high-wage workers using public transit. After Phase 1 rail is completed and workers are allowed to reoptimize their location, work, and mode-choice decisions, I find public transit mode share increases to 20.4% for low-wage workers and to 10.2% for high-wage workers. I find a larger effect for Phase 2 rail, which provides a rail option for a larger share of commuting routes. After workers reoptimize according to Phase 2 rail (Scenario 5) I find low and high wage worker transit mode shares rise to 24.3% and 15.2% respectively. Comparing Scenario 1 to 5, I find that the overall public transit mode share increases from 13.3% to 19.8%,

Figure 5: Changes in Aggregate Outcomes



The graphs show the progression of rail's effect on three outcomes. Scenario 1 corresponds to the pre-rail period while Scenario 5 corresponds to the full rail system with endogenous worker choices. Full scenario descriptions are provided in Figure 4.

a 49% increase. The majority of the improvement (68%) occurs on account of Phase

2 rail. Phase 2 also attracts relatively more high-wage workers to public transit, as the Phase 2 stations serve more routes with high preference parameters for high-wage workers.

The Scenario 1 solution shows that the average one-way commute time for a low-wage worker is 22.4 minutes while the average for a high-wage worker is 20.1 minutes. The changes in commute times are summarized in Figure 5B. The introduction of Phase 1 rail lowers average commute times as workers who used transit along the rail route benefit from shorter commuting times (Scenario 2). The majority of initial commute time benefits accrue to low-wage workers, who are currently the primary users of transit on O'ahu, particularly along the routes served by Phase 1 rail. After the opening of Phase 1 rail, island-wide average low-wage commute time falls to 21.8 minutes (a 2.6% reduction) while high-wage average commute time falls to 20.0 (a 0.6% reduction). Once endogenous worker choices are allowed, all of the commuting time gains are erased. The primary mechanism that causes rail to result in higher commute times is that transit is a slower mode of transportation, even after the improvements attributable to rail. The increase in public transit mode share (Figure 5A) translates to a rise in average commute time. As a second order effect, the allocation of rail represents a local amenity to the neighborhoods with rail stations, pushing up local housing costs. Because the location decision of low-wage workers are sensitive to rents, this causes some low-wage workers to leave the areas for locations of lower housing costs. Low housing cost areas tend to be more peripheral, and often include longer commutes.

In Scenario 4, with the introduction of the full Phase 2 rail line, both low and high wage commute times fall significantly. The relative effect on high-wage workers is larger in Phase 2 because the location of the new stations align more closely with existing high-wage commute flows. After I allow for full endogenous sorting (Scenario 5) I find commute times rise again. In the final equilibrium, I find that average commute time across all O'ahu workers increases by 1.0% (or .20 minutes) compared to a scenario where rail was never built. The introduction of transit systems are often meant to reduce commuting times. It is important to note that when endogenous worker choices are considered, the improvement of public transit infrastructure is likely to raise the average commuting time across the labor market. While I do not account for potentially improved traffic conditions due to mode switching away from private vehicles, in the long run, induced demand suggests that the time savings will be negligible (Duranton and Turner, 2011).

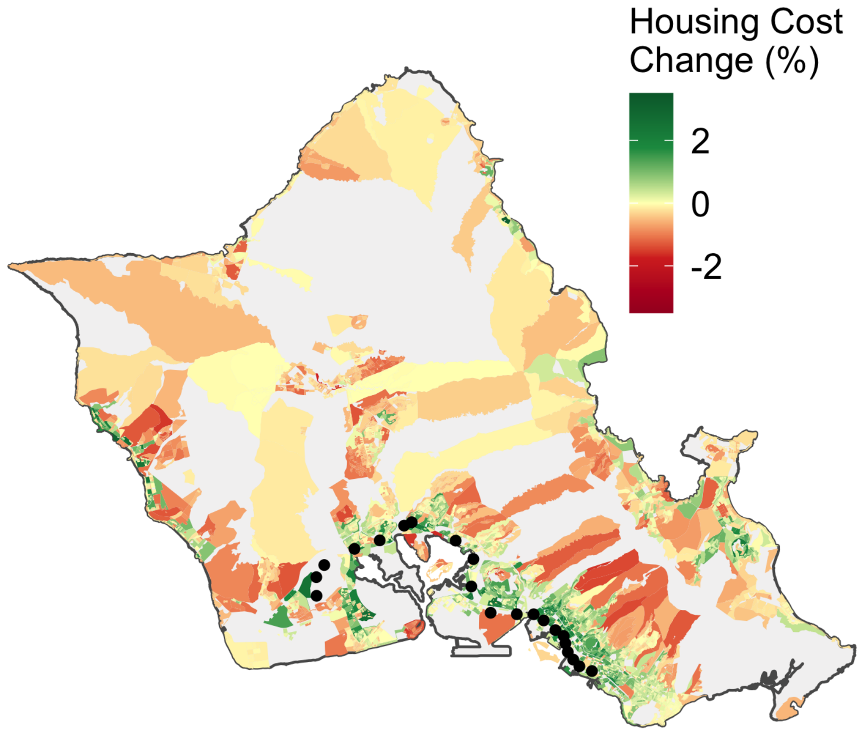
Figure 5C summarizes the effects on the share of workers who are employed. High commuting costs are a disincentive to employment. The provision of rail allows a worker to access more employment opportunities for a given amount of commuting costs. Depending on worker idiosyncratic preferences across home location, work location, and mode, the reduced commuting costs will push marginal workers into employment. Across all workers, I find the full Phase 2 rail system increases the employment rate by 1.5 percentage points from 66.4% to 67.8%. The effect among low-wage workers is a 1.6 point increase whereas the effect on high-wage workers is 1.2 points. The smaller effect among high-wage workers is a result of public transit playing a smaller role in the commute choice of many high-wage workers.

Figure 6 displays the block level estimated changes in housing cost (p_j) between Scenario 1 and 5. I estimate significant increases in housing costs for blocks near to the new rail stations. The block experiencing the largest increase in housing costs sees an increase of 3.2%, while the largest decrease experienced is 1.7%. The cost increases near to stations are largely offset by rent decreases in neighborhoods far from stations, which become comparatively less desirable. Some interesting substitution patterns emerge from the model. For example, I find price increases in the O'ahu neighborhood of Kailua, located in the north-east section of the island, despite Kailua being far from rail. Routes originating from Kailua have high preference parameters among high-wage workers. The substitution pattern is consistent with high-wage workers moving away from rail neighborhoods towards Kailua. High-wage workers are less likely to use rail but would still need to pay the higher housing costs associated with increased neighborhood demand. Therefore rail may push out high-wage workers and cause them to select alternative neighborhoods which match their preferences.

7 Conclusion

I estimate the effects of O'ahu's forthcoming rail system through a structural approach. I show that modeling endogenous worker decisions is key to estimating the aggregate effects of the system. By directly modeling worker behavior I am able to provide realistic estimates of aggregate rail impacts. While a common motivation for constructing transit improvements is to reduce commute times, I find that the O'ahu system is likely to marginally increase the average time spent commuting by a worker on O'ahu. However, this is due to the system's success in shifting a meaningful number of workers away from private vehicle commuting to public transit commuting. Furthermore, the option of reasonably fast and affordable public transit encourages some

Figure 6: Estimated Changes in Local Housing Costs



The map shows the predicted housing cost effects of the rail system at the block level. Prices generally increase near rail stations and fall elsewhere. Areas with no housing are shown in grey. Rail stations are shown as black dots.

workers to enter the labor force. I estimate the full rail system will increase O'ahu's employment rate by 1.5 percentage points.

One limitation of the model is the assumption of a "closed city." The creation of a valuable public amenity is likely to make workers from outside of O'ahu marginally more likely to move to O'ahu, which may fuel further rent increases around stations and have other second order effects. Modeling workers as independent agents is also a limitation as many workers are in dual-earner households and face a more complex location optimization problem. A complementary policy to rail on O'ahu has been an attempt to generate new housing near rail stations through zoning changes that encourage Transit Oriented Development. I do not model endogenous housing supply responses, and consider this process to be separate from the impacts of rail. Despite these limitations, I believe the paper provides realistic estimates for the probable effects of rail. All of the paper's main results are driven by endogenous worker choices, which highlights the importance of urban structural modeling in evaluating urban transit projects. This paper contributes to the literature on discrete neighborhood choice modeling and related studies on transportation infrastructure evaluation.

I analyze a data set with richer spatial variation than has been attempted in any prior related works. Census block level analysis allows for the model to capture extremely local impacts of rail. Workers are rarely willing to walk significant distances to reach rail. Many studies assume pedestrian catchment areas extend only about 0.5 miles from a station (Guerra et al., 2012). Therefore, the use of larger geographic units will be unable to accurately capture commuter incentives. The combination of multiple worker types, explicit modeling of transportation costs, and the use of route by worker-type level preference parameters provides a unique modeling approach that may be helpful for research in other settings.

References

Ahlfeldt, G. M., Redding, S. J., Sturm, D. M., and Wolf, N. (2015). The economics of density: Evidence from the Berlin wall. *Econometrica*, 83(6):2127–2189.

Alonso, W. (1964). *Location and land use. Toward a general theory of land rent.* Cambridge. Harvard University Press.

American Automobile Association (2021). *Your Driving Costs*.

Anas, A. (1981). The estimation of multinomial logit models of joint location and travel mode choice from aggregated data. *Journal of Regional Science*, 21(2):223–242.

Andersson, F., Haltiwanger, J. C., Kutzbach, M. J., Pollakowski, H. O., and Weinberg, D. H. (2018). Job displacement and the duration of joblessness: The role of spatial mismatch. *Review of Economics and Statistics*, 100(2):203–218.

Bayer, P., Ferreira, F., and McMillan, R. (2007). A unified framework for measuring preferences for schools and neighborhoods. *Journal of Political Economy*, 115(4):588–638.

Bayer, P. and McMillan, R. (2012). Tiebout sorting and neighborhood stratification. *Journal of Public Economics*, 96(11-12):1129–1143.

Bayer, P., McMillan, R., and Rueben, K. (2004). An equilibrium model of sorting in an urban housing market. Technical report, National Bureau of Economic Research.

Bayer, P. and Timmins, C. (2005). On the equilibrium properties of locational sorting models. *Journal of Urban Economics*, 57(3):462–477.

Brooks, L. and Liscow, Z. (2022). Infrastructure costs. *American Economic Journal: Applied Economics*.

Chernoff, A. and Craig, A. N. (2022). Distributional and housing price effects from public transit investment: Evidence from Vancouver. *International Economic Review*, 63(1):475–509.

Davis, M. A. and Ortalo-Magné, F. (2011). Household expenditures, wages, rents. *Review of Economic Dynamics*, 14(2):248–261.

Duranton, G. and Turner, M. A. (2011). The fundamental law of road congestion: Evidence from us cities. *American Economic Review*, 101(6):2616–52.

Epple, D. and Sieg, H. (1999). Estimating equilibrium models of local jurisdictions. *Journal of Political Economy*, 107(4):645–681.

Fujita, M. and Ogawa, H. (1982). Multiple equilibria and structural transition of non-monocentric urban configurations. *Regional Science and Urban Economics*, 12(2):161–196.

Guerra, E., Cervero, R., and Tischler, D. (2012). Half-mile circle: Does it best represent transit station catchments? *Transportation Research Record*, 2276(1):101–109.

Gupta, A., Van Nieuwerburgh, S., and Kontokosta, C. (2022). Take the Q train: Value capture of public infrastructure projects. *Journal of Urban Economics*, 129:103422.

Holzer, H. J., Quigley, J. M., and Raphael, S. (2003). Public transit and the spatial distribution of minority employment: Evidence from a natural experiment. *Journal of Policy Analysis and Management*, 22(3):415–441.

Kain, J. F. (1968). Housing segregation, negro employment, and metropolitan decentralization. *The Quarterly Journal of Economics*, 82(2):175–197.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics*, pages 105–142.

Mills, E. S. (1967). An aggregative model of resource allocation in a metropolitan area. *The American Economic Review*, 57(2):197–210.

Muth, R. F. (1969). Cities and housing; the spatial pattern of urban residential land use.

Sanchez, T. W. (1999). The connection between public transit and employment: the cases of Portland and Atlanta. *Journal of the American Planning Association*, 65(3):284–296.

Severen, C. (2019). Commuting, labor, and housing market effects of mass transportation: Welfare and identification. *The Review of Economics and Statistics*, pages 1–99.

Sieg, H., Smith, V. K., Banzhaf, H. S., and Walsh, R. (2004). Estimating the general equilibrium benefits of large changes in spatially delineated public goods. *International Economic Review*, 45(4):1047–1077.

Small, K. A., Winston, C., and Yan, J. (2005). Uncovering the distribution of motorists' preferences for travel time and reliability. *Econometrica*, 73(4):1367–1382.

Tsivanidis, N. (2018). The aggregate and distributional effects of urban transit infrastructure: Evidence from Bogota's Transmilenio. *Working Paper, University of Chicago Booth School of Business*.

Tyndall, J. (2017). Waiting for the R train: Public transportation and employment. *Urban Studies*, 54(2):520–537.

Tyndall, J. (2018). Bus quality improvements and local commuter mode share. *Transportation Research Part A: Policy and Practice*, 113:173–183.

Tyndall, J. (2021). The local labour market effects of light rail transit. *Journal of Urban Economics*, 124:103350.